New York State Department of Health

Service Oriented Architecture: A Practical Approach to Data Exchange, Integration and Linkage for Environmental Public Health Tracking Network

Linh H. Le, M.D., Ph.D., MPH

Research Scientist, BHNSM, NYSDOH

New York State Department of Health

Preface

- SOA is just another ploy to give money to contractor and IBM
- Web services are a mess!
- Simple? Even ADA is simpler
- Just a bunch of competing specs for software vendor to make more money
- These specs are denser than plutonium
- SOA is just a letter away from SOB

New York State Department of Health

What is SOA

- Collection of services that communicate with each other
- Evolution of distributed computing based on the request/reply design paradigm for synchronous and asynchronous applications
- Self-contained and independent of the context or state of the other service
- Business logic or individual functions are modularized and presented as services for consumer/client applications

New York State Department of Health

Key Characteristics of SOA

- Service interface is self-described in XML documents - Web Services Description Language (WSDL)
- Communication between services is based on messages defined via XML Schema
- Services are maintained by a registry that acts as a directory listing - Universal Description, Definition, and Integration (UDDI)
- Each SOA service has a quality of service (QoS) associated with it

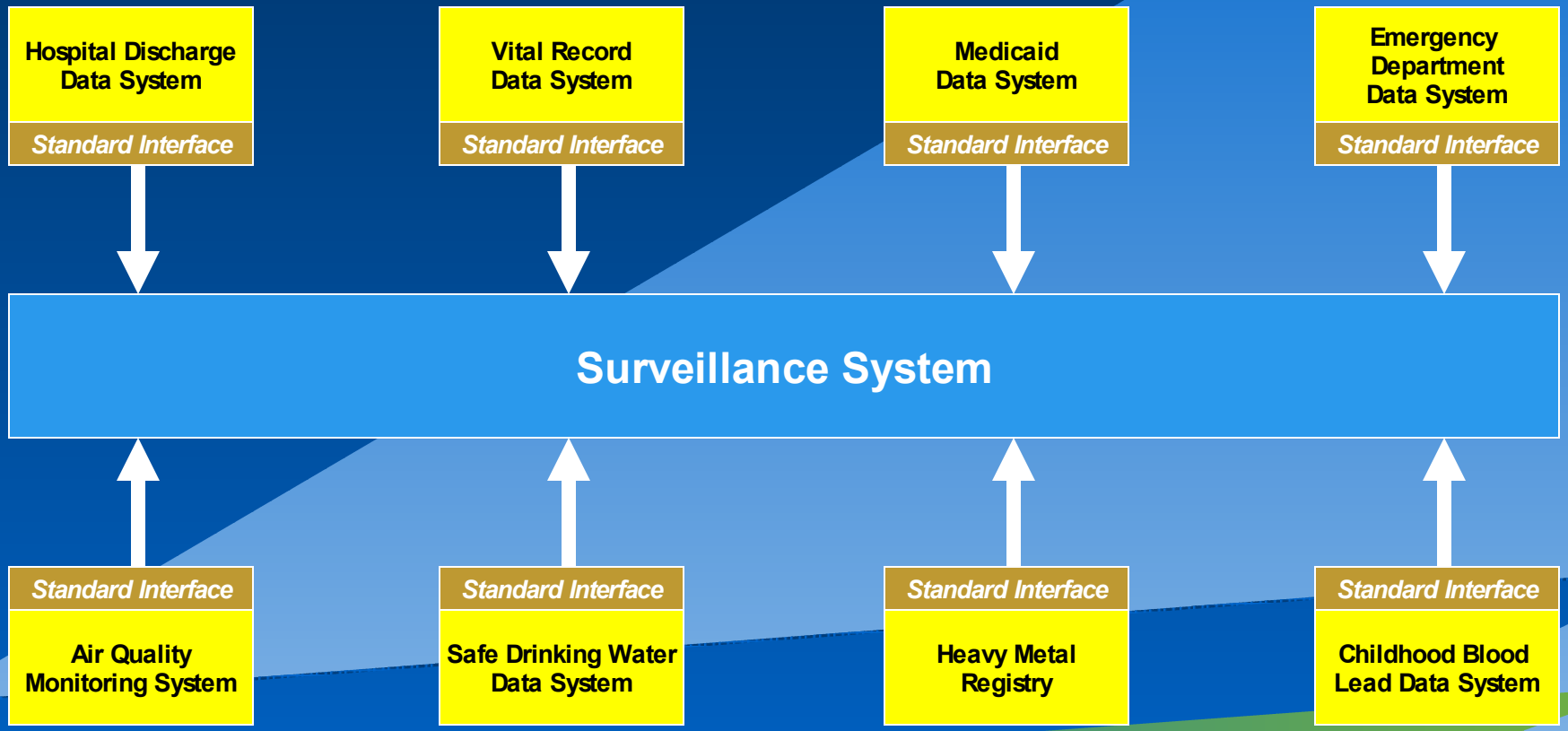
New York State Department of Health

Why SOA

- Support for Programmatic (Business) Process Optimization
- Platform-neutral approach to accessing services
- Better interoperability as more and more vendors support more and more Web services specifications
- Need to respond quickly to changes with agility
- Need to leverage existing infrastructure to address new programmatic requirements
- Need to support new channels of communication with public health partners

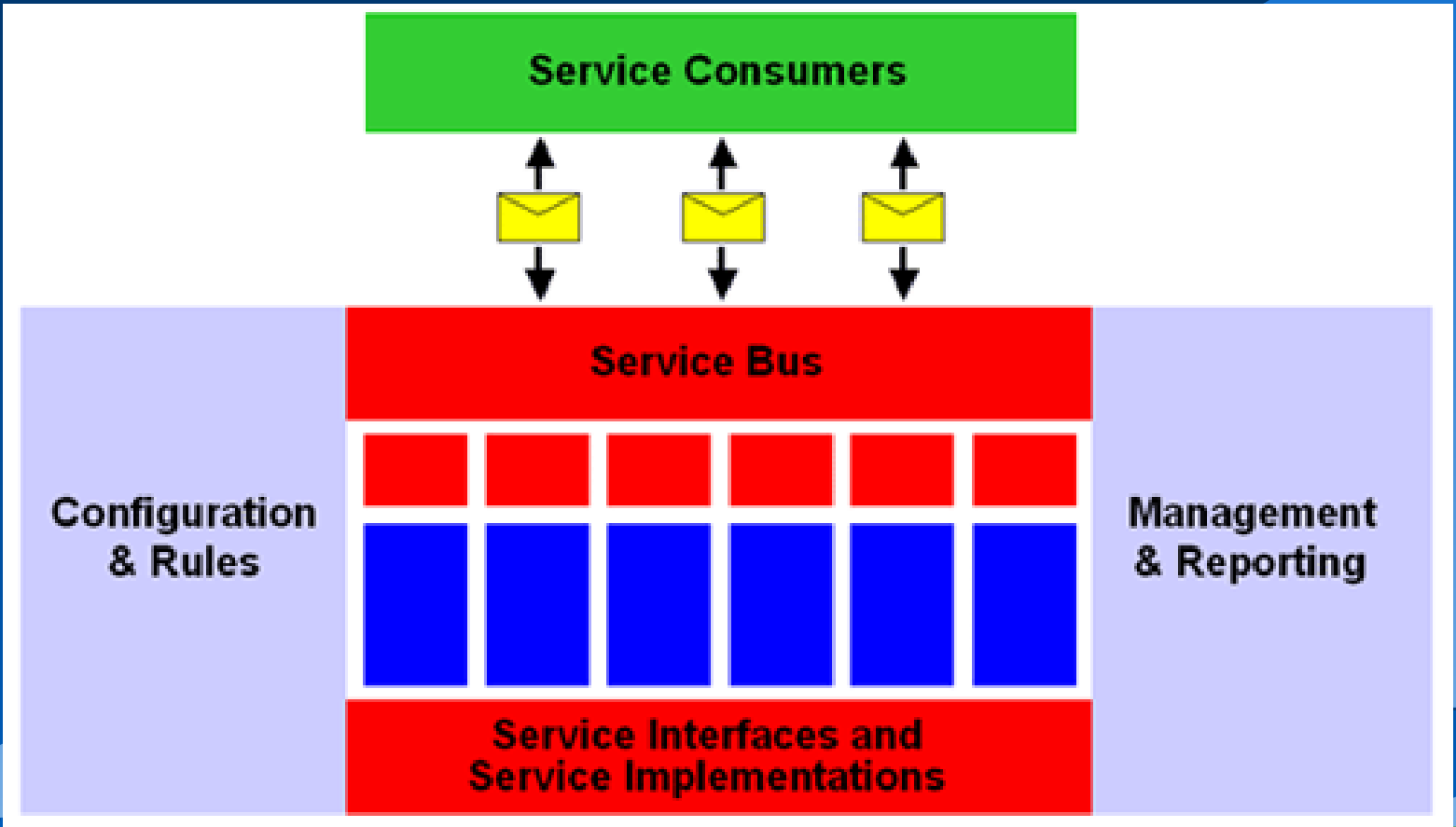
New York State Department of Health

Example of a SOA based Surveillance System



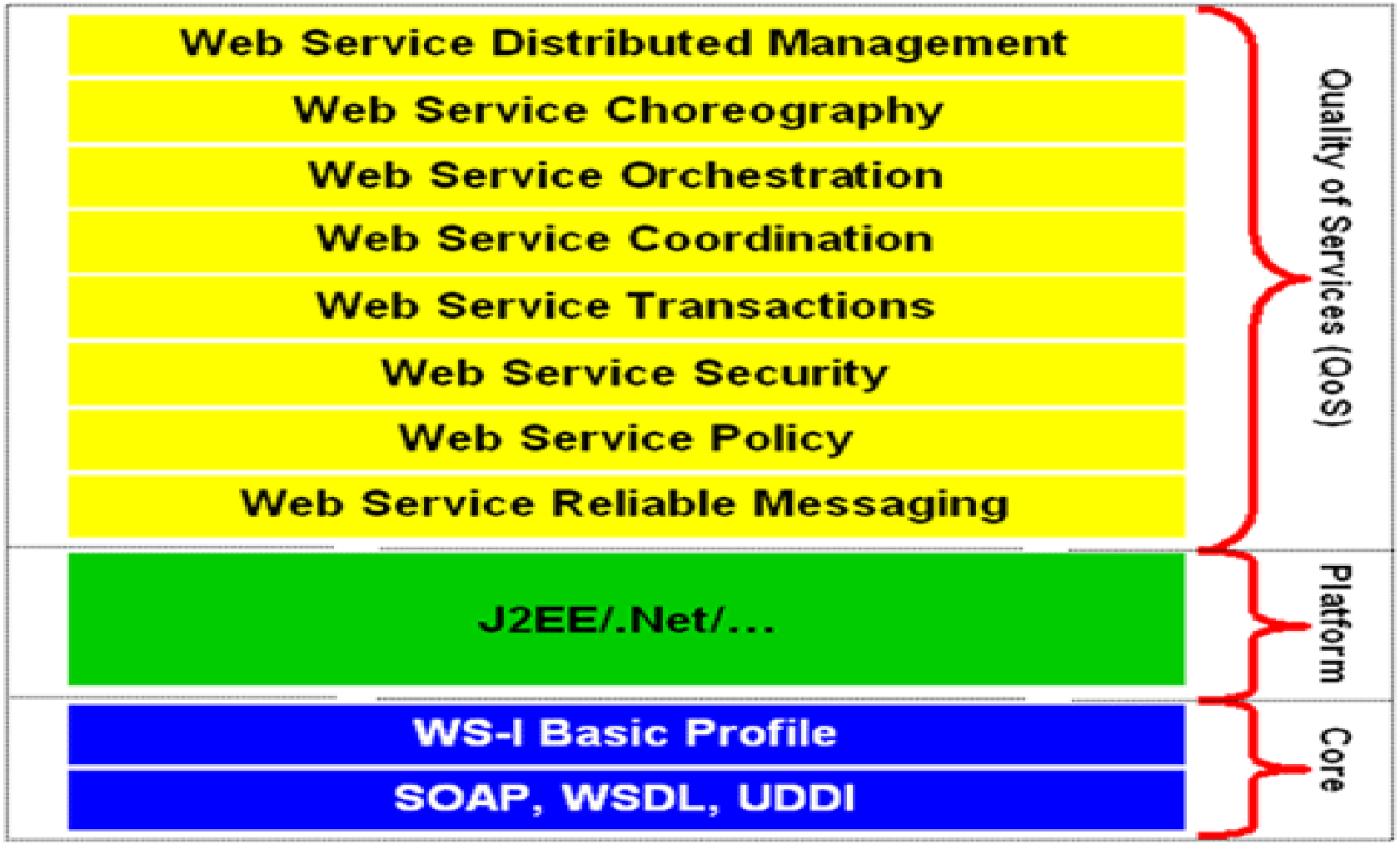
New York State Department of Health

Service Architecture



New York State Department of Health

SOA Infrastructure



New York State Department of Health

SOAP, WSDL, UDDI

- WSDL is used to describe the service
- UDDI is used to register and look up the services
- SOAP is used as a transport layer to send messages between service consumer and service provider
- While SOAP is the default mechanism for Web services, alternative technologies accomplish other types of bindings for a service
- SOAP 1.1 and SOAP with Attachments specifications were integrated into the ebXML Messaging Specification

New York State Department of Health

SOA is not Web services

- General confusion about the relationship between SOA and Web services
- SOA is a software design principle
- Web services are about technology specifications
- Web services is one of the ways to implement SOA
- Web services is a standards-based, XML centric realization of SOA

New York State Department of Health

Deriving Web Services from SOA

- Everyone knows roughly what a "web service" is, but there is no universally accepted definition
- It is generally accepted that a web service is a SOA with at least the following additional constraints
  - *Interfaces must be based on Internet protocols such as HTTP, FTP, and SMTP*
  - *Except for binary data attachment, messages must be in XML*
- There are two main styles of Web services: SOAP web services and REST web services

New York State Department of Health

SOAP Web services

- SOAP web service is the most common and marketed form of web service in the industry
- SOAP acts like an envelope that carries its contents
- There are two flavors of SOAP web services, SOAP RPC and document-centric SOAP web service
- SOAP RPC web services are not SOA; document-centric SOAP web services are SOA

New York State Department of Health

REST Web Services

- Representational State Transfer (REST) web services is an SOA based on the concept of "resource (URI)“ and have become very popular
- Interfaces are limited to HTTP with the four well-defined verbs: GET, POST, PUT, and DELETE
- Most messages are in XML, confined by a schema written in a schema language such as XML Schema from W3C or RELAX NG.
- Simple messages can be encoded with URL encoding
- Service and service providers must be resources while a consumer can be a resource

New York State Department of Health

Advantages of REST

- REST is easy to implement and has many highly desirable architectural properties: scalability, performance, security, reliability, and extensibility
- Require little infrastructure support apart from standard HTTP and XML processing technologies

New York State Department of Health

Hidden SOA Challenges

- A still-emerging infrastructure
- Service governance
- Change to the definition of data can affect any service that handles the data
- SOA tie the services together into a business process, effectively coupling those services through the business process and the data
- It is harder to change the interface of a service with a large number of consumer
- More integration can lead more technical complexity

New York State Department of Health

Why SOA for EPHT Network

- A shift in information technology from Data Orientation to Service Orientation
- Building new infrastructure is almost impossible
- Support for Resources Sharing in true distributed computing environment
- Ability to go beyond data transport
- Ability to integrate different components of the network such as portal, gateway
- Support for multi-dimensional data linkage

New York State Department of Health

SOA and Data Exchange

- Web services provide system-to-system, via defined formats and protocols, and are capable of processing large amounts of data across the Internet
- Numerous specifications are being developed to address advanced requirements such as security, reliability, and transactions
- EPA Exchange Network is an example of using Web services for exchanging data with partners

New York State Department of Health

SOA and Data Integration

- Traditional data integration provides adequate business information (data replication) without up-to-the-minute access of information
- In many cases, the data is weeks, even months, old, and the data mart or data warehouse is updated through antiquated batch, extract-aggregate-and-load, processes
- SOA can support real-time data communication but is there an actual need for it ?

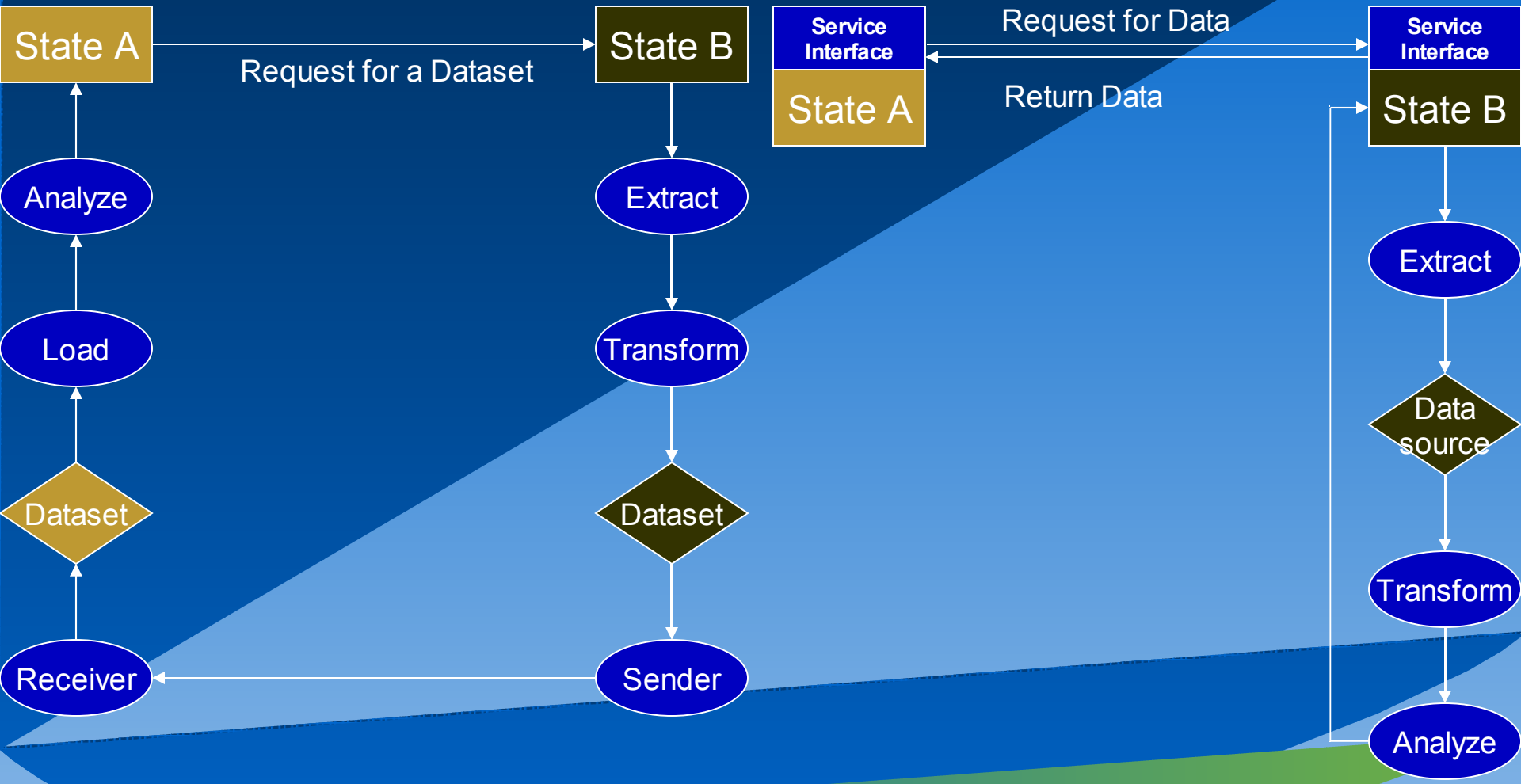
New York State Department of Health

Why SOA for Data Integration

- Today is Tuesday, August 15, 2006 and things are changing
- SOA will also lead to the rise of real-time data warehouse solutions, or could replace the notion of data warehousing all together
- SOA can provide services to abstract both operational and aggregated data, in some cases leveraging aggregated data without having to replicate and change the data, but instead do it through abstraction layers

New York State Department of Health

Traditional vs SOA Data Integration



New York State Department of Health

Data Linkage

- Allows multi-dimensional data linkage
- Preserve data confidentiality and privacy
- Provide on-demand data linkage
- Easier to implement and enforce security
- Support one to one as well as one to many data linkage
- Reduce cost and time
- Provide add-on data modeling and processing services

New York State Department of Health

SOA Data Linkage Example

- An environmental agency analyzes air monitoring data and predict exceed level of Ozone in certain areas
- Via a SOA service, a request is sent to health department SOA service for asthma hospitalization
- Request parameters include locations with high ozone level based on city state, zip code, or a combination of any of these things, latitude-longitude, ID of air modeling grid, or a search radius in miles, type of requested data, ICD code of the health outcome, format of the returned output
- https://api.epht.commerce.health.state.ny.us/HealthService/V1/sparcsData?id=aft07338&icd=493&gridID=03601&type=count&age=0+5&sex=male

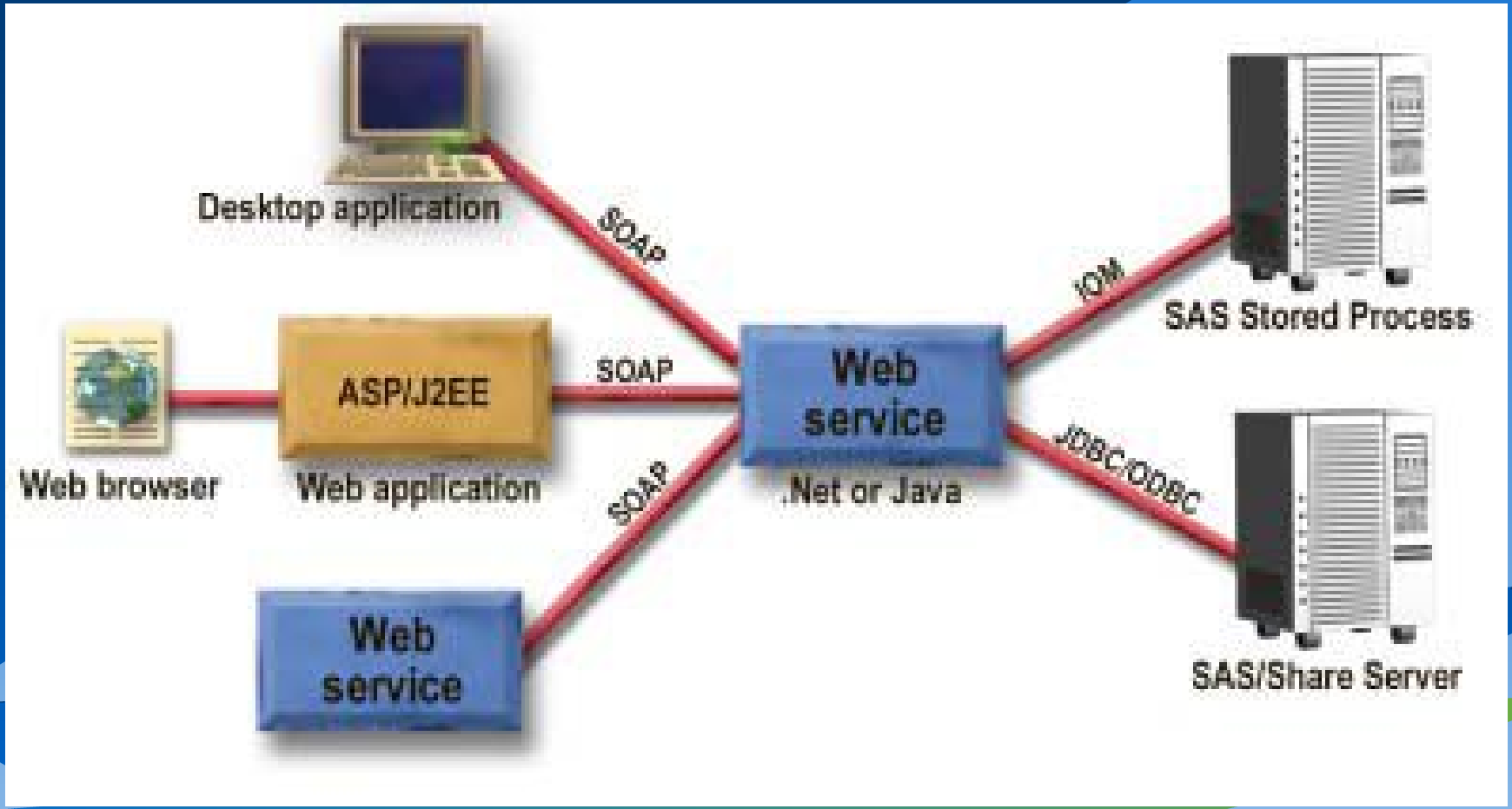
New York State Department of Health

What is next

- Based on the parameter of the request, Health Department SOA interface make a query to the hospital discharge database for all asthma cases (ICD = 493), male sex, age group from 0 to 5
- A geocoding service is called to geocode the records, then a geoprocessing service is called to perform a point in layer analysis to assign the grid ID to each record
- Finally, the number of cases that is in gridID 03601 returned in a response to the environmental agency SOA service

New York State Department of Health

Using SAS with .NET and Java Web Services infrastructures



New York State Department of Health

Conclusions

- SOA or Web service is not the magical answer for everything
- While technical issues are being addressed, there is a number of programmatic, organizational and cultural barriers to SOA
- SOA will require a higher level of collaboration and communication between technical and programmatic staff especially data owners